Can. J. Math., Vol. XXVIII, No. 1, 1976, pp. 135–140

CODING ERGODIC PROCESSES TO APPROXIMATE BERNOULLI PROCESSES

A. DEL JUNCO

1. Introduction. In [**1**] Ornstein defined a metric $\bar{d}$ on processes which, for processes (P, τ) and (Q, σ) with equal numbers of atoms, measures how closely the motions of P and Q under τ and σ, respectively, imitate each other. If we think of (P, τ) and (Q, σ) as stationary stochastic processes, and we assume (P, τ) and (Q, σ) are ergodic, then $\bar{d}((P, \tau)(Q, \sigma)) < \alpha$ says that with probability one a printout from (P, τ) can be changed on a set of integers with density less than α to obtain a printout from (Q, σ). This metric was first introduced (implicitly) by Ornstein in [**3**] to characterize Bernoulli processes (i.e. processes isomorphic to independent processes), but as he shows in [**1**] it seems to be a natural way to compare stationary stochastic processes. From this point of view, however, the restriction to processes with equal numbers of atoms seems artificial, so in this paper we consider the obvious extension of $\bar{d}$ to a metric d which allows comparison of processes with different numbers of atoms.

Since independent processes (and, more generally, Bernoulli processes) are particularly nice it seems natural to ask to what extent arbitrary processes can be approximated by them. It is known [**1**, p. 54] that the class of Bernoulli processes with a given number of atoms is closed in the metric $\bar{d}$, and it follows easily that the class of all Bernoulli processes is closed in the metric d. Thus for a given process (Q, σ), $d((Q, \sigma), B) > 0$ where B denotes the class of Bernoulli processes. However we shall show that if (Q, σ) is ergodic and (P, τ) is any Bernoulli process with the same entropy as (Q, σ), then one can find a process (Q', σ') isomorphic to (Q, σ) which is arbitrarily close to (P, τ) in d-distance. Moreover there is a bound on the number of atoms in Q' which is independent of ϵ. From the point of view of stationary stochastic processes the significance of (Q, σ) and (Q', σ') being isomorphic is that (Q, σ) can be approximated arbitrarily well in d-distance by finite codings of (Q', σ'), and vice-versa. For a discussion of coding see [**1**].

Note that in purely ergodic theoretic terms this result is just Theorem 4 below. The proof of Theorem 4 is an application of some of the techniques introduced by Ornstein in [**2**] together with his characterization of Bernoulli processes as those that are finitely determined [**3**]. The reader is referred to [**4**] for a relaxed exposition of these matters.

Theorem 4 is taken from my thesis written at the University of Toronto

Received February 5, 1975.

https://doi.org/10.4153/CJM-1976-016-2 Published online by Cambridge University Press

under the supervision of M. A. Akcoglu. I would like to thank Prof. Akcoglu for his encouragement and many helpful conversations on this result.

2. Definitions and notation. If τ is a measure preserving automorphism of a finite measure space (X, F, μ) and P is a finite measurable partition of X, the pair (P, τ) is called a process. We denote by $P^{m,n} (m \leqq n)$ the partition $\bigvee_{i=m}^{n} \tau^i P$ and by $P^{-\infty,\infty}$ the σ-field generated by the $\tau^i P$, $i \in Z$. Another partition P' of X is called a generator for (P, τ) if $P'^{-\infty,\infty} = P^{-\infty,\infty}$. If $P = \{P_\gamma : \gamma \in \Gamma\}$ is a partition indexed by a set Γ, dist P is a measure on Γ defined by

$$\text{dist } P\{\gamma\} = \frac{\mu(P_\gamma)}{\mu(X)}.$$

Thus dist $P^{m,n}$ is a measure on $\Gamma^{[m,n]}$ ($[m, n] = \{m, m + 1, \ldots, n\}$). The P n-name of a point $x \in X$ is a sequence ξ in $\Gamma^{-(n-1),0}$ defined by

$$\xi(-i) = \gamma \text{ if } \tau^i x \in P_\gamma,\ 0 \leqq i \leqq n - 1.$$

Two processes (P, τ) and (Q, σ) indexed by the same set are called equivalent ($(P, \tau) \sim (Q, \sigma)$) if dist $P^{-n,n} =$ dist $Q^{-n,n}$ for all $n \in N$. If P is a partition of X and $F \subset X$, $P|_F$ denotes the partition $\{p \cap F : p \in P\}$. If $x \in \Gamma^{[m,n]}$ and $r \leqq m - n + 1$ we define a measure $\text{fr}_x{}^r$ on $\Gamma^{-(r-1),0}$ by

$$\text{fr}_x{}^r\{y\} = \frac{1}{m - n + 2 - r} \text{card} \{i: (x(i), \ldots, x(i + r - 1)) = (y - (r - 1)), \ldots, y(0))\}$$

for $y \in \Gamma^{-(r-1),0}$, that is $\text{fr}_x{}^r\{y\}$ is the frequency with which the string y appears as a string of consecutive symbols in x.

If P and Q are partitions of a probability space (X, F, μ) with possibly different numbers of atoms we define

$$\rho(P, Q) = \min \mu \bigcup_{\gamma \in \Gamma} (P_\gamma \Delta Q_\gamma)$$

where the minimum is over all indexings of P and Q by a common set Γ, allowing P_γ or Q_γ to be empty for some γ's. For processes (P, τ) and (Q, σ) on possibly different spaces we say $d((P, \tau), (Q, \sigma)) < \alpha$ if there exists a space Y with an ergodic automorphism U and partitions $\bar{P}$ and $\bar{Q}$ such that $(\bar{P}, U) \sim (P, \tau)$, $(\bar{Q}, U) \sim (Q, \sigma)$, and $\rho(\bar{P}, \bar{Q}) < \alpha$. We set

$$d((P, \tau), (Q, \sigma)) = \inf \{\alpha : d((P, \tau), (Q, \sigma)) < \alpha\}.$$

This corresponds to the first definition of the metric $\bar{d}$ given in [**1**, Section 2], except that for $\bar{d}$ one requires partitions with the same numbers of atoms and a given correspondence between their atoms.

We also need the analogous extension of the distribution metric on partitions. If (P, τ) and (Q, σ) are processes we say $\Delta(P^{-n,0}, Q^{-n,0}) < \alpha$ if for some indexing of P and Q by a common set Γ

$$\frac{1}{2} \sum_{x \in \Gamma^{[-n,0]}} |\text{dist } P^{-n,0}\{x\} - \text{dist } Q^{-n,0}\{x\}| < \alpha.$$

(Notice that we do not allow arbitrary correspondences between the atoms of $P^{-n,0}$ and $Q^{-n,0}$—the correspondence must be determined by a fixed correspondence between P and Q.) We set

$$\Delta(P^{-n,0}, Q^{-n,0}) = \inf \{\alpha : \Delta(P^{-n,0}, Q^{-n,0}) < \alpha\}.$$

We will also write

$$\text{dist } P^{-n,0} \overset{\alpha}{\sim} \text{dist } Q^{-n,0} \text{ if } \Delta(P^{-n,0}, Q^{-n,0}) < \alpha.$$

The definition of finitely determined given in [**1**, Appendix 2], is in terms of $\bar{d}$ and the distribution metric but we get an equivalent definition if we use d instead of $\bar{d}$ and Δ instead of the distribution metric as long as we put a bound on the number of atoms in the processes involved. More precisely, (P, τ) is finitely determined if for each $\epsilon > 0$ and $r \in N$, there exist $\delta > 0$ and $n \in N$ such that if $(\bar{P}, \bar{\tau})$ is any process with fewer than r atoms satisfying $\Delta(\bar{P}^{-n,0}, P^{-n,0}) < \delta$ and $h((\bar{P}, \bar{\tau}), (P, \tau)) < \delta$ then $d((\bar{P}, \bar{\tau}), (P, \tau)) < \epsilon$. The proof of this is straightforward. The bound on the number of atoms is required to keep the entropy under control.

Before proceeding to the theorem we state without proof a simple lemma that will be needed.

LEMMA 3. *For every $\epsilon > 0$ there is a $\delta > 0$ such that if P is a partition on a probability space X and $X^* \subset X$, $\mu(X^*) > 1 - \delta$, then*

$$||\text{dist } P|_{F} - \text{dist } P|| < \epsilon.$$

THEOREM 4. *Given any ergodic process (Q, σ), a Bernoulli process (P, τ) such that $h(P, \tau) = h(Q, \sigma)$ and $\epsilon > 0$, there exists a generator Q' for the process (Q, σ) such that $d((Q', \sigma), (P, \tau) < \epsilon$. The number of atoms in Q' has a bound independent of ϵ.*

Proof. Suppose P has b atoms, Q has c atoms. (P, τ) is Bernoulli, hence finitely determined [**5**, Theorem 11.3] so there exist an $r \in N$ and a $\delta > 0$ such that for any ergodic process $(\bar{P}, \bar{\tau})$ with fewer than $4b + c$ atoms.

$$\text{(i)} \left.\begin{array}{c} \Delta(\bar{P}^{-(r-1),0}, P^{-(r-1),0}) < \delta \\ \text{and} \\ |h(\bar{P}, \bar{\tau}) - h(P, \tau)| < \delta \end{array}\right\} \Rightarrow d((\bar{P}, \bar{\tau}), (P, \tau)) < \epsilon.$$

We now construct an auxiliary process $(\hat{P}, \hat{\tau})$ with slightly higher entropy. Let (B, T) be an independent process on a probability space Y such that B has two atoms, one of measure $\delta/2r$, the other of measure $1 - \delta/2r$. Let $(\bar{P}, \hat{\tau}) = (P \times B, \tau \times T)$. It is clear that $\Delta(P \times Y, P \times B) < \delta/2r$ whence

$$\Delta((P \times Y)^{-(r-1),0}, (P \times B)^{-(r-1),0}) < \delta/2.$$

https://doi.org/10.4153/CJM-1976-016-2 Published online by Cambridge University Press

Since dist $(P \times Y)^{-(r-1),0} = \text{dist } P^{-(r-1),0}$ it follows that

(ii) $\Delta(\hat{P}^{-(r-1),0}, P^{-(r-1),0}) < \delta/2$.

We also have

(iii) $\hat{h} = h(\hat{P}, \hat{\tau}) = h(P, \tau) + h(B, T) > h(P, \tau) = h$.

Notice that $(\hat{P}, \hat{\tau})$ is ergodic.

Let $\eta > 0$. Let $n > r/\eta$ be a positive integer to be further specified in the course of this proof. Assume $\hat{P}$ is indexed by Ω. The individual ergodic theorem and the Shannon-McMillan-Breiman theorem together imply that if n is sufficiently large there is a set $A_n \subset \Omega^{[-(n-1),0]}$ such that

(iv) $(\text{dist } \hat{P}^{-(n-1),0})A_n > 1 - \eta$

(v) $\text{fr}_x^{\,r} \overset{\eta}{\sim} \text{dist } \hat{P}^{-(r-1),0}$ for $x \in A_n$

(vi) $e^{-(\hat{h}+\eta)n} < \text{dist } \hat{P}^{-(n-1),0}\{x\} < e^{-(\hat{h}-\eta)n}$ for $x \in A_n$.

(iv) and (vi) imply

(vii) $\text{card } A_n > e^{(\hat{h}-\eta)n}(1 - \eta) > e^{(\hat{h}-2\eta)n}$

if n is sufficiently large.

Suppose (Q, σ) acts on the probability space (X, F, μ). The Shannon-McMillan-Breiman theorem implies that if n is sufficiently large there is a set $G_n \subset X$ which is a union of atoms of $Q^{-(n-1),0}$ such that

(viii) $\mu(G_n) > 1 - \eta$

(ix) $e^{-(h+\eta)n} < \mu(E) < e^{-(h-\eta)n}$ for all $E \in Q^{-(n-1),0}$ such that $E \subset G_n$.

(ix) implies

(x) $\text{card } (Q^{-(n-1),0}|G_n) < e^{(h+\eta)n}$.

By Rohlin's theorem [**5**, Theorem 8.1] there is a set $F' \in Q^{-\infty,\infty}$ (note that $h(Q, \sigma) > 0$ so $Q^{-\infty,\infty}$ is non-atomic), such that $\sigma^i(F')$ $0 \leqq i \leqq n - 1$ are mutually disjoint and

$$\mu\left(\bigcup_{i=0}^{n-1} \sigma^i F'\right) > 1 - \eta.$$

Since

$$\mu\left[\left(\bigcup_{i=0}^{n-1} \sigma^i F'\right) \cap G_n\right] > 1 - 2\eta$$

there must be an i such that

$$\mu(\sigma^i F' \cap G_n) > \frac{1 - 2\eta}{n}.$$

Let $F = \sigma^i F' \cap G_n$. Then we have that $\sigma^i F$, $0 \leqq i \leqq n - 1$, are mutually disjoint, and

$$\mu \bigcup_{i=0}^{n-1} \sigma^i F > 1 - 2\eta.$$

https://doi.org/10.4153/CJM-1976-016-2 Published online by Cambridge University Press

Now since $F \subset G_n$, (x) implies that card $(Q^{-(n-1),0}|_F) < e^{(h+\eta)n}$. Comparing this with (vii) we see that if $\eta < (\hat{h} - h)/4$, then card $A_n >$ card $(Q^{-(n-1),0}|_F)$. Let f be any one to one map from $Q^{-(n-1),0}|_F$ to A_n. This assignment of $(\hat{P}, \hat{\tau})$ n-names to points in F defines a partition $\tilde{P}$ of $X^* = \bigcup_{i=0}^{n-1} \sigma^i F$, indexed by Ω, in the following way. For each atom E of $Q^{-(n-1),0}|_F$ and each i, $0 \leqq i \leqq n-1$, $\sigma^i E \subset \tilde{P}_\omega$ if $f(E)(-i) = \omega$. In other words, $\tilde{P}$ is the unique partition of X^* such that for each $x \in E \in Q^{-(n-1),0}|_F$, the $\tilde{P}$ n-name of x is just $f(E)$. Let $\bar{P}$ be the partition of X whose atoms are the atoms of $\tilde{P} \vee \{\sigma^{n-1}F, X^* - \sigma^{n-1}F\}$ together with the atoms of $Q|_{X-X^*}$. Note that card $\bar{P} \leqq 4b + c$.

Let us show that $\bar{P}$ is a generator for (Q, σ). Since different atoms of $Q^{-(n-1),0}$ have different $\bar{P}$ n-names it is easy to see that $\bar{P}^{-(n-1),0}|_{X^*}$ refines $Q|_{X^*}$. Since we also have $\bar{P}|_{X-X^*} = Q|_{X-X^*}$ and X^* is a union of atoms of $\bar{P}$ it follows that $\bar{P}^{-(n-1),0}$ refines Q. Thus $\bar{P}$ is a generator for (Q, σ) and in particular $h(\bar{P}, \sigma) = h(Q, \sigma)$.

It remains only to show that $d((\bar{P}, \sigma), (P, \tau)) < \epsilon$. Let

$$X^{**} = \bigcup_{i=0}^{n-r} \sigma^i F.$$

From the definition of $\tilde{P}$ it is clear that for each atom E of $Q^{-(n-1),0}|_F$

(xi) $\text{dist } \tilde{P}^{-(r-1),0}\Big|_{\bigcup_{i=0}^{n-r}\sigma^i E} = \text{fr}^r_{f(E)}$.

(xi) and (v) imply

(xii) $\text{dist } \tilde{P}^{-(r-1),0}|_{X^{**}} \overset{\eta}{\sim} \text{dist } \hat{P}^{-(r-1),0}$.

Since $r < \eta n$ and $\mu(X^*) > 1 - 2\eta$ we have

$$\mu(X^{**}) > (1-\eta)(1-2\eta) > 1 - 3\eta.$$

Thus Lemma 3 implies that, given η_1, if η is sufficiently small, then

$$\begin{aligned} \text{dist } \bar{P}^{-(r-1),0} &\overset{\eta_1}{\sim} \text{dist } \bar{P}^{-(r-1),0}|_{X^{**}} \\ &= \text{dist } \tilde{P}^{-(r-1),0}|_{X^{**}} \\ &\overset{\eta}{\sim} \text{dist } \hat{P}^{-(r-1),0} \quad \text{by (xii)} \\ &\overset{\delta/2}{\sim} \text{dist } P^{-(r-1),0} \quad \text{by (ii)} \end{aligned}$$

Thus if η_1 is so small that $\eta + \eta_1 < \delta/2$ we have

$$\text{dist } \bar{P}^{-(r-1),0} \overset{\delta}{\sim} \text{dist } P^{-(r-1),0}.$$

Since we also have $h(\bar{P}, \sigma) = h(Q, \sigma) = h(P, \tau)$ and card $\bar{P} \leqq 4b + c$, (i) concludes the proof.

References

1. D. S. Ornstein, *An application of ergodic theory to probability theory*, Ann. Probability *1*, (1973), 43–65.
2. ——— *Bernoulli shifts with the same entropy are isomorphic*, Advances in Math. *4* (1970), 337–352.
3. ——— *Factors of Bernoulli shifts are Bernoulli shifts*, Advances in Math. *5* (1970), 349–364.
4. P. C. Shields, *The theory of Bernoulli shifts* (The University of Chicago Press, 1973).
5. M. Smorodinsky, *Ergodic theory, entropy*, Springer Lecture Notes in Mathematics, No. 214 (1971).

University of Toronto,
Toronto, Ontario